

# **RMC Research Corporation**

*A Resource Management Corporation Subsidiary*

Report UR-257

## FIELD TEST AND EVALUATION OF A FUNCTIONAL ASSESSMENT SYSTEM FOR ADULTS NEEDING LONG-TERM CARE

### Executive Summary

Michael T. Errecart  
Donald H. Strope

November 1977

Prepared for

Social and Rehabilitation Service  
Department of Health, Education, and Welfare

Under

Contract No. SRS-74-53

REPORTS

RA

644

.6

E77

1977

*Home Office:*

*7910 Woodmont Avenue, Bethesda, Maryland 20014*

*Other Offices:*

*Mountain View, California; Los Angeles, California*

*Honolulu, Hawaii; Portsmouth, New Hampshire*



RA  
644.6  
.E77  
1977

RMC Report  
UR-257

FIELD TEST AND EVALUATION OF A FUNCTIONAL ASSESSMENT  
SYSTEM FOR ADULTS NEEDING LONG-TERM CARE

Executive Summary

Michael T. Errecart  
Donald H. Strope

November 1977

This report is made pursuant to Contract No. SRS-74-53. The amount charged to the Department of Health, Education, and Welfare for the work resulting in this report (inclusive of the amounts so charged for any prior report submitted under this contract) is \$310,159. The names of the persons employed or retained by the contractor with management or professional responsibility for such works, or for the content of the report, are as follows: Michael T. Errecart and Donald H. Strope.

Prepared for

Social and Rehabilitation Service  
Department of Health, Education, and Welfare

Under  
Contract No. SRS-74-53



CMC Library  
C2-07-13  
7500 Security Blvd.  
Baltimore, MD 21246



## ACKNOWLEDGEMENTS

The RMC study team wishes to acknowledge the continual assistance and suggestions of the SRS Project Officer, Edward Neuschler. We also appreciate the efficient and professional work done by RMC's subcontractor --National Certified Interviewers, Incorporated, of Chicago--in conducting the field interviews.

Further, the field work portion of our study could not have been accomplished without the cooperation of the private citizens and the institutional staffs who volunteered to participate in the survey.



## FOREWORD

The Social and Rehabilitation Service (SRS) developed a study design intended to assess the functional status of large groups of people with possible long-term care problems. At the heart of the design was a "multidimensional functional assessment questionnaire" and classification system. The questionnaire and system had been created by the Older American Resources and Services program of the Duke University Center for the Study of Aging and Human Development.

As part of its evaluation of the study design, SRS, along with the Administration on Aging (AoA), entered into a contract with RMC Research Corporation to test the validity, reliability, and usefulness of the questionnaire and functional classification system. The Administration on Aging also provided project support. This volume summarizes the various analyses conducted by RMC and presents conclusions regarding the questionnaire and system, as well as recommendations for further research.

Volume 2 presents the complete analysis results, conclusions, and recommendations. A companion volume details RMC's evaluation strategies for testing the Duke System and RMC's field survey design and procedures for carrying out the analysis plan. In two earlier tasks, RMC prepared draft reports on the design of a national longitudinal survey. The purpose of those tasks was to provide preliminary planning for a future national survey using the Duke system or some variation of it.



## CONTENTS

Acknowledgements . . . . .	ii
Foreword . . . . .	iii
Background of the Study . . . . .	1
Research Context . . . . .	2
Functional Classification System . . . . .	3
Field Test . . . . .	4
Results . . . . .	8



## EXECUTIVE SUMMARY

### BACKGROUND OF THE STUDY

In recent years, there has been considerable debate over the costs and appropriateness of long-term care services provided to impaired people. This debate has tended to focus on the institutionalization/de-institutionalization issue. All sides can agree, however, that determining what is appropriate care (i.e., the most effective care at a given cost) is not a simple matter. Consideration must be given to the relative effectiveness, for each type and degree of impairment, of various service packages delivered in various settings. And, in each case, cost estimates must be made that reflect total social costs--not just government outlays. Current and previous surveys of long-term care have not covered the entire long-term care universe in a consistent fashion, and none of them have included a method of classifying clients by type and degree of impairment, examined the full range of possible services, or permitted assessment of total social costs.

Despite the volume of data available about long-term care patients and facilities, little of it is useful for evaluating government policy with respect to long-term care services. Also, the more meager data available on the non-institutionalized impaired population are not generally useful for policy purposes. It is quite clear that, before policy can be rationally made in this area, a great deal of empirical evidence must be obtained on the appropriateness and cost of care for individuals with different types and levels of impairments.



## RESEARCH CONTEXT

The Social Rehabilitation Service (SRS) of DHEW has developed a research design to learn more about the functional status of impaired individuals and the costs, supply, and use of long-term care for the impaired elderly and for the emotionally disturbed, mentally retarded, chronically ill, physically handicapped, and traumatically injured of all ages. The purpose of the SRS design is to enable DHEW to test various hypotheses about the appropriateness of care and total social costs.

As a first step in implementing this design, three developmental research projects were jointly funded by SRS and the Administration on Aging. These projects are or were being conducted by:

- (1) Duke University, Center for the Study of Aging and Human Development, Older American Resources and Services Program;
- (2) University of Rochester, Department of Pediatrics, Psychodiagnostic Laboratory; and
- (3) RMC Research Corporation.

These simultaneous and coordinated research efforts will provide the procedures and instruments for establishing a data base to assess alternative policy and program intervention approaches. The essential element and major focus of the program has been the development, refinement, and testing of a system for classifying the functional impairments of large population groups.

Duke University has developed a functional classification system and applied it to a population of elderly people residing in either their own homes or institutions. The system uses a questionnaire administered by nonclinicians to classify individuals by their level of functional impairment.

The University of Rochester has adapted the Duke University classification system to be appropriate for the long-term care population that is less than 18 years in age. That instrument was field tested with 400 impaired children and adolescents living at home or in institutions.



RMC's project has field tested the validity, reliability, and usefulness of the Duke University functional classification system. The field test applied the Duke instrument to 703 adults in three geographically disparate regions of the continental United States. Specifically, the target population included both institutionalized and non-institutionalized impaired elderly and emotionally disturbed, mentally retarded, chronically ill, physically handicapped, or traumatically injured adults.

## **FUNCTIONAL CLASSIFICATION SYSTEM**

The classification system developed by Duke University uses functional level, rather than diagnosis, as the common yardstick. Although diagnosis and other specific labeling of impairments may be more useful when assessing individuals, the unifying concept of the functional level is uniquely suitable for classifying and tracking a population over time and across several dimensions.

### **Multidimensional Assessment**

Unlike other large-scale survey efforts that are limited to narrow categories of impairment or diagnosis, the Duke system collects information on all major functional areas of a person's daily life, including:

- (1) physical health,
- (2) mental health,
- (3) capacity for activities of daily living (ADL),
- (4) social resources, and
- (5) economic resources.

Information in each of these areas is collected through personal interviews given by nonclinical interviewers using a detailed questionnaire. In each dimension, both factual and observational data as well as subjective perceptions are gathered. Information is obtained on a comparable basis for the full spectrum of major long-term impairments for both institutionalized and non-institutionalized individuals.

### **Functional Rating**

Through empirical work carried out by various clinicians and methodologists, a common system for determining functional level has been



defined by the Duke University staff. Based on the information collected during the interview, a rating of one through six is assigned in each of the five functional areas, as follows:

- 1 = outstanding.
- 2 = OK, adequate,
- 3 = mild impairment,
- 4 = moderate impairment,
- 5 = severe impairment, and
- 6 = complete impairment.

To further clarify and anchor the meaning of a rating for each functional area, there is a brief descriptive paragraph. For example, the definition of "mildly physically impaired" is: "Has only minor illness and/or disabilities that might benefit from medical treatment or corrective measures."

#### FIELD TEST

##### Surveys Conducted

Three surveys were used to gather the field test data, and each was conducted in both households and institutions. The first survey involved 703 personal interviews from 352 people in households and 351 in institutions. The purpose of the main interview was to obtain information using the full household and institution questionnaires so Duke clinicians could rate each respondent according to the Duke classification system. The main survey was conducted by RMC's subcontractor, National Certified Interviewers, Incorporated, of Chicago.

The number of people interviewed in the main survey, by location and living arrangement was:

<u>County</u>	<u>Number of Personal Interviews</u>		
	<u>Household</u>	<u>Institution</u>	<u>Total</u>
Los Angeles, California	151	201	352
Carver/Hennepin, Minnesota	75	100	175
Washington, Mississippi	126	50	176
Total	352	351	703



The second survey reinterviewed a subsample of 60 respondents: 30 in households and 30 in institutions. The purpose of the reinterviews was to check the stability of selected questions in the questionnaire. A briefer version of the full questionnaire was used in the reinterviews to minimize respondent burden. To reduce survey costs, the reinterviews were conducted only in California and Minnesota. The reinterviews were also conducted by National Certified Interviewers.

The number of reinterviews conducted in each sampled county was:

<u>County</u>	<u>Number of Reinterviews</u>		
	<u>Household</u>	<u>Institution</u>	<u>Total</u>
Los Angeles, California	20	20	40
Carver/Hennepin, Minnesota	10	10	20
Total	30	30	60

The third survey was a reinterview with another subsample of 121 persons. These reinterviews were conducted by clinicians who used their own interview and examination approaches and then rated the respondents according to the Duke classification system. In particular, the Duke questionnaire was not used. Physicians were asked to make physical health ratings, psychologists to make mental health ratings, and social workers to rate social resources, economic resources, and the functional capability to perform the activities of daily living.

The purpose of these reinterviews was to provide clinical ratings for comparison with the Duke ratings in a test of the external validity of the rating procedure. To reduce costs and logistical problems, the clinical interviews/examinations were restricted to California and Minnesota. In addition, resources did not permit the use of multiple assessments by different clinicians. Thus, we have no measures of the reliability of the external validators.

The criteria for selecting the clinicians to conduct the external validation interviews were that the clinicians should be highly trained and should have clinical experience in the functional areas for which they would be providing ratings.

Recruiting experienced clinicians to participate in the study within the funds available required numerous contacts with professional associations, heads of medical staffs, mental health centers, social work



agencies, and various individuals recommended to us. The individuals recruited not only met the criteria, but they were also interested in the research aims of this study.

The number of clinical interviews in each county was:

<u>County</u>	<u>Number of Clinical Interviews</u> <sup>a</sup>		
	<u>Household</u>	<u>Institution</u>	<u>Total</u>
Los Angeles, California	57	65	122
Carver/Hennepin, Minnesota	30	34	64
Total	87	99	186 <sup>b</sup>

- a. Two interviews could not be classified and do not appear in the table.
- b. It should be noted that physicians only evaluated physical health, psychologists only evaluated mental health, and social workers evaluated economic resources, social resources, and activities of daily living. Thus, any one respondent may have been interviewed up to three times, which accounts for the number of interviews substantially exceeding the number of respondents.

#### Selection of Field Sites

The selection of the three field test States reflected variations in geographic region, general level of State resources, and State interest in social programming. Counties in the three States were selected to represent variations in population size, urban-rural setting, socioeconomic factors, age distribution, public expenditures for health and welfare services, and types of health-related residential facilities.

The choice of the county as the primary sampling unit reflected the fact that the interviewer work load per county approximated that expected for a national survey. As in the selection of the States, the sample counties were chosen with certainty rather than randomly because generalizations about the national long-term care population would not be attempted.

Los Angeles County, with a population of over seven million, was selected as representative of large metropolitan areas. Two counties from the other States were chosen to represent smaller population strata: Washington County, Mississippi (70,000-75,000 population stratum), and Carver County, Minnesota (less than 30,000 population stratum). These three counties were used in both the household and institution surveys. However, because Carver County had few long-term care institutions, we also selected nearby Hennepin County, Minnesota, for the institution survey.



In the sample counties, census areas (enumeration districts/block groups) were randomly selected and segmented. Within those sampling units, all households were screened until enough eligible respondents had been identified. Screening was accomplished through a brief questionnaire intended to ascertain whether anyone in the household had, or potentially had, a long-term health care problem.

Because the screening questions could have been biased in the way they selected respondents, we developed and tested two screening alternatives. One approach (Form A) asked about the limitations in a person's ability to function normally, such as whether he or she had to stay in bed most of the time or needed a lot of help from others in everyday activities. The second approach (Form B) asked about specific long-term health problems, such as deafness, eye trouble, diabetes, cancer.

Somewhat different selection procedures were used to identify the institutionalized respondents. First, institutions located in the sample counties were selected from the Master Facility Inventory maintained by the National Center for Health Statistics. Then, at the institution, RMC staff members randomly selected a sample of patients from the institutional population. The administrator of the institution was asked to name a staff member who knew the patient well enough to answer questions about him or her.

#### Data Sources

Up to three sources of information were tapped for each interview. The primary source of information was the respondent. But, before the interview took place, the respondent was informed that we intended to ask questions of another informed person in a separate interview. If the proposed respondent understood and agreed to the second interview, then both interviews were held. Otherwise, neither interview was conducted. The third source of information was the interviewer, who filled out a short questionnaire concerning the length of the interview and some opinions about the reliability and situation of the respondent.

#### Ratings

The original rating concept allowed nonclinical interviewers to determine the functional status of the respondent. However, during the



survey clearance process, the Office of Management and Budget objected to lay persons making what it considered clinical judgments. As a result, SRS changed the process so the interviewers only conducted the interviews -- the functional ratings were assigned by clinicians based solely on the questionnaires. The revised process had the advantage of having the ratings made by people trained in assessing health and functional status and the disadvantage of ratings being made by people who did not benefit from the cues inherent in face-to-face situations.

The completed questionnaires were sent to Duke University to be rated. Duke used six experienced clinical staff members from its Older American Resources and Services (OARS) program in the Center for the Study of Aging and Human Development. All 703 household and institution questionnaires were split fairly evenly among six clinicians and were rated on the Duke scale of one to six in each of the five functional areas: physical health, mental health, activities of daily living (ADL), social resources, and economic resources.

Additional ratings were obtained to assess inter-rater reliability (i.e., the degree to which different raters agree in their assessments of the same respondent). This was done by having each of the six Duke raters separately rate the same 25 questionnaires. Twelve of the questionnaires were from the household survey and 13 from the institution survey. These questionnaires were randomly selected and can be presumed to be representative.

## RESULTS

### Response Rates

Nonresponse was not a significant problem in the personally administered applications of the Duke instrument. Respondents did not appear reluctant to answer questions, except for certain economic resource items dealing with specific income levels and assets. For those questions, nonresponse rates of up to 15 percent were encountered.

### Response Stability

Questions dealing with mental health, social resources, and general physical health were extremely unstable. In about 30 percent of the



resurveyed cases, the second responses on those questions differed from the initial responses.

A second group of questions exhibited instability rates between 10 and 15 percent. This group included many of the activities of daily living (ADL) questions and certain economic resources questions.

In our opinion, the questions in the highly unstable group should be closely scrutinized for deletion in any future application of the questionnaire. The use of such questions could result in highly unstable ratings. For example, the social rating is highly correlated (above .8) to variable S-9 (whether someone would care for the subject, if necessary), which is highly unstable. We have no direct evidence of this instability because the resurvey was not rated. Nevertheless, we do know that the raters disagreed most often over the social rating and that the Duke-assigned social ratings appeared to be significantly biased with respect to the external ratings.

Further analyses should be conducted of the unstable variables before deletion, however. The analysis in this report merely looked at the incidence of discrepancy; it did not consider the size of discrepancy. Such an analysis would provide a better characterization of the discrepancies, especially in the case of variables measured at ordinal and interval levels.

#### Inter-Rater Reliability

The statistical tests of the raters' behavior could not detect any significant differences between the raters in the assignment of ratings to respondents. This is not to say that the raters produced identical ratings--they did not--but rather that no systematic differences were detected.

Typically, the raters tended to agree on an individual rating in about 64 percent to 88 percent of the ratings assigned. Furthermore, instances of ratings deviating by more than one point from the consensus rating were rare (less than 3 percent in all dimensions).

It should be noted that the set of raters for this particular study was probably unique. The raters all had a long involvement with the questionnaire and its previous applications, and they worked together



for a long time. Consequently, there could have been an enhanced rating commonality that might not be characteristic of other groups of raters.

### External Validity

Of fundamental importance to institutions considering using the Duke system is the extent to which the ratings based on the questionnaire data agree with the ratings assigned through more familiar procedures; i.e., professional judgments by competent clinicians based on personal contacts.

The worth of the approach tested in this study depends on the precision and type of inferences one would like to draw. For example, the questionnaire approach can produce ratings that diverge by as much as four points from professional judgments. Thus, as a tool for individual classification, the Duke procedures are not sufficiently accurate. But, if one wanted a conservative estimate of the population, the approach could be used.

We found that the questionnaire-based ratings tended to be more conservative than the external validator ratings; i.e., questionnaire-based ratings tended to indicate greater impairment of the subject. The extent of the bias varied from area to area. Economic and ADL ratings showed little bias, mental ratings a small amount, and social ratings a significant conservative bias. Inferences about the physical ratings are difficult to make because the range of the ratings in the sample was essentially only three points. Nevertheless, a conservative bias was indicated. In each rating area, the ratings assigned under both methods were positively correlated. In other words, the methods tended to agree regarding the statuses of the subjects.

In the aggregate, it appears that the questionnaire procedure produces a reasonable profile of the respondents in the economic and ADL dimensions. In the other three dimensions, the procedure tends to significantly overstate the impairment level of the population. Whether it is possible to adjust for this overstatement is a subject for further research, as the data on hand are not adequate for a thorough analysis. There are two major deficiencies:



- (1) In order to evaluate differences at the extremes of the rating scale, the sample ought to contain approximately the same number of people at each functional level in each dimension; this is not the case with the present data.
- (2) In order to evaluate the indeterminacy of the validator's judgments, multiple validators ought to be used; limited resources precluded such an approach in the current study.

### Internal Validity

The internal validity and dimensionality tests uncovered a significant number of variables that might be deleted from the instrument. Further, they identified variables closely related to the ratings.

In our opinion, the ratings, as currently defined, do in fact indicate five independent, if not orthogonal, rating dimensions. We found evidence of clusters of variables forming around each of the five dimensions.

This is not to say that all five dimensions should be kept in a future questionnaire. In particular, our analysis showed the social rating to be essentially an alternate expression of a particular question (S-9, whether someone would care for the subject, if necessary), which was one of the most unstable variables in the analysis. Further, we found that the Duke raters exhibited the most disagreement over the social ratings. The disagreement between the Duke raters and the external validators was also greatest in this dimension. In addition, few variables other than S-9 revealed any clustering tendency with the social rating.

Two other dimensions need to be examined closely--mental health and physical health.

The mental dimension is distinct from ADL, but it is clearly a related concept. The mental variables, especially the subunits of M-3 (a 15-item scale), are somewhat unstable. Nevertheless, good inter-rater reliability was evidenced as well as only a small bias with respect to the external validators.

The physical dimension is clearly distinct from the other dimensions. Nevertheless, it is a difficult dimension to analyze because the questions are structured into many subunits inquiring about specific conditions. We strongly question the level of detail in this section of the instrument.



Clusters of variables were identified in all areas of the questionnaire. The strongest clusters generally depicted redundancies in the instrument or possibilities that were too rare to be significant categories.

### Screening Forms

Two forms (A and B) were used to evaluate whether a potential respondent to the household survey should be included in the study. The possibility exists that these systems for excluding households resulted in the selection of respondents with significantly different functional statuses.

For each dimension, we compared the distribution of ratings assigned to individuals screened under each form. We found that the Form A population did not differ from the Form B population in the social, economic, or physical dimensions. We found the Form A population to be significantly more impaired than the Form B population in the ADL and mental categories. This result was not unexpected, since Form A focused on whether the respondent was limited in his or her ability to perform normal, everyday activities, whereas Form B inquired about specific conditions, many of which would not immediately cause limitation of activity.

The best screening form depends on the population of interest. We can offer no guidance other than to indicate that the choice of screening procedures is an important consideration that can significantly affect the results of any application of the instrument.





CMS LIBRARY



3 8095 00007940 6